

# Hierarchical Clustering for Automatic File Classification: A Study on Scalable Approaches

**Start Date:** 05-10-2024

**End Date:** 25-11-2024

**Team Members:**

Kameswaran R - 2020239007

Viswaranjani S - 2020239023

Vigneshwarann Vijayakumar - 2020242023

Date	Activity	Progress
12-10-2024	Initiated research on clustering techniques for unlabeled datasets.	Studied key clustering metrics, including Silhouette Score and Davies-Bouldin Index, for evaluating clustering performance.
20-10-2024	Developed feature extraction methods tailored to different file types.	Implemented different techniques for feature extraction for different type of files.
02-11-2024	Focused on clustering evaluation and visualization.	Computed linkage matrices using agglomerative clustering methods.
10-11-2024	Incorporated advanced metrics and refined the clustering process.	Implemented computation of intra-cluster and inter-cluster distances.
24-11-2024	Finalized the project and documented findings.	Conducted a comprehensive evaluation of the system using a mixed file dataset.